

University of Groningen

Considering Horn's parallel analysis from a random matrix theory point of view

Saccenti, Edoardo; Timmerman, Marieke E.

Published in:
Psychometrika

DOI:
[10.1007/s11336-016-9515-z](https://doi.org/10.1007/s11336-016-9515-z)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Final author's version (accepted by publisher, after peer review)

Publication date:
2017

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Saccenti, E., & Timmerman, M. E. (2017). Considering Horn's parallel analysis from a random matrix theory point of view. *Psychometrika*, 82(1), 186-209. <https://doi.org/10.1007/s11336-016-9515-z>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Psychometrika

CONSIDERING HORN'S PARALLEL ANALYSIS FROM A RANDOM MATRIX THEORY POINT OF VIEW --Manuscript Draft--

Manuscript Number:	PMET-D-14-00122R3
Full Title:	CONSIDERING HORN'S PARALLEL ANALYSIS FROM A RANDOM MATRIX THEORY POINT OF VIEW
Abstract:	<p>Horn's parallel analysis is a widely used method for assessing the number of principal components and common factors. We discuss the theoretical foundations of parallel analysis for principal components based on a covariance matrix by making use of arguments from random matrix theory. In particular, we show that i) for the first component, parallel analysis is an inferential method, equivalent to the Tracy-Widom test, ii) its use to test high order eigenvalues is equivalent to the use of the joint distribution of the eigenvalues, and thus should be discouraged, and iii) a formal test for higher order components can be obtained based on a Tracy-Widom approximation. We illustrate the performance of the two testing procedures using simulated data generated under both a principal component model and a common factors model. For the principal component model, the Tracy-Widom test performs consistently in all conditions, while parallel analysis shows unpredictable behavior for higher order components. For the common factor model, including major and minor factors, both procedures are heuristic approaches, with variable performance. We conclude that the Tracy-Widom procedure is preferred over parallel analysis for statistically testing the number of principal components based on a covariance matrix.</p>
Keywords:	covariance matrix; Tracy-Widom distribution; dimensionality; principal component analysis; common factor analysis; number of principal components; number of common factors; major factors; minor factors
Article Type:	Original Research
Section/Category:	Theory and Methods (T&M)
Corresponding Author:	Edoardo Saccenti Wageningen University NETHERLANDS
Corresponding Author's Institution:	Wageningen University
Corresponding Author E-Mail:	esaccenti@gmail.com
Order of Authors:	Edoardo Saccenti Marieke Timmerman
Author Comments:	NA
Response to Reviewers:	As discussed with the Editorial office the code is going to be share through our website. It is already available at semantics.systemsbiology.nl as stated in the manuscript.
Additional Information:	
Question	Response
If your article is accepted for publication or a reviewer requests electronic supplementary material, are you willing to share any related data and/or code to accompany your article? In the interest of transparency and reproducibility of results, Psychometrika strongly encourages authors to include with	Yes

published papers, where possible, electronic supplementary materials which will be published online on SpringerLink with the article.	
<p>Please specify the type of supplementary material you will be able to provide. Data and code should be clearly documented to allow replication and verification of the results presented in the final version of the paper.</p> <p>as follow-up to "If your article is accepted for publication or a reviewer requests electronic supplementary material, are you willing to share any related data and/or code to accompany your article? In the interest of transparency and reproducibility of results, Psychometika strongly encourages authors to include with published papers, where possible, electronic supplementary materials which will be published online on SpringerLink with the article. "</p>	Code

CONSIDERING HORN'S PARALLEL ANALYSIS FROM A RANDOM MATRIX THEORY POINT OF VIEW

Edoardo Saccenti^{1,*} and Marieke E. Timmerman^{2,#}

¹Wageningen University, ²Univerisity of Groningen

Corresponding Author

*Edoardo Saccenti: Laboratory of Systems and Synthetic Biology, Wageningen University and Research Center, Dreijenplein 10, 6703 HB, Wageningen, The Netherlands. Email: esaccenti@gmail.com; Tel: +31 (0)317 482018; Fax: +31 (0) 3174 83829

Co-Corresponding Author

Marieke E Timmerman: Department Psychometrics & Statistics , Univerisity of Groningen, Grote Kruisstraat 2/1, 9712 TS, Groningen, The Netherlands. Email: m.e.timmerman@rug.nl; Tel: +31 (0)50 363 6255



CONSIDERING HORN'S PARALLEL ANALYSIS FROM A RANDOM
MATRIX THEORY POINT OF VIEW

February 17, 2016

CONSIDERING HORN'S PARALLEL ANALYSIS FROM A RANDOM MATRIX THEORY POINT OF VIEW

Abstract

Horn's parallel analysis is a widely used method for assessing the number of principal components and common factors. We discuss the theoretical foundations of parallel analysis for principal components based on a covariance matrix by making use of arguments from random matrix theory. In particular, we show that *i*) for the first component, parallel analysis is an inferential method equivalent to the Tracy-Widom test, *ii*) its use to test high order eigenvalues is equivalent to the use of the joint distribution of the eigenvalues, and thus should be discouraged, and *iii*) a formal test for higher order components can be obtained based on a Tracy-Widom approximation. We illustrate the performance of the two testing procedures using simulated data generated under both a principal component model and a common factors model. For the principal component model, the Tracy-Widom test performs consistently in all conditions, while parallel analysis shows unpredictable behavior for higher order components. For the common factor model, including major and minor factors, both procedures are heuristic approaches, with variable performance. We conclude that the Tracy-Widom procedure is preferred over parallel analysis for statistically testing the number of principal components based on a covariance matrix.

Key words: covariance matrix; Tracy-Widom distribution; dimensionality; principal component analysis; common factor analysis; number of principal components; number of common factors; major factors; minor factors

Introduction

Determining the dimensionality underlying sample data is an important step in scale construction and in psychological theory building (Hattie, 1985). Preceding a principal component analysis (PCA) or common factor analysis (CFA), the dimensionality is typically indicated using an empirical criterion. Many of those criteria are based on the behavior of the eigenvalues of the sample covariance matrix or sample correlation matrix. Examples are the eigenvalues-greater-than-one criterion (Guttman, 1954), Cattell's scree-plot (Cattell, 1966) and Horn's parallel analysis (Horn, 1965; Glorfeld, 1995). Parallel analysis appears to be among the most recommended methods (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Hayton, Allen, & Scarpello, 2004; Thompson, 2004) to indicate the number of principal components and common factors.

The basic idea of Horn's parallel analysis is to compare the k^{th} eigenvalue of the sample covariance or correlation matrix with the sampling distribution of the k^{th} eigenvalue, which is obtained via Monte-Carlo simulation from random, independent data. The sampling distribution provides a reference distribution under the null hypothesis of independence. Under independence, the dimensionality is zero, and thus one would retain zero components. A k^{th} eigenvalue larger than the 95th percentile of the k^{th} eigenvalue sampling distribution (as proposed by Glorfeld (1995)) indicates to retain the k^{th} component. For the first eigenvalue, this procedure can be viewed as a statistical test. For the remaining eigenvalues (*i.e.*, $k > 1$) this interpretation is not straightforward, because of the inherent dependencies between successive eigenvalues. Buja and Eyübgölu (1992) still interpreted the procedure for $k > 1$ in terms of a statistical test, albeit with reduced power. Other investigators explicitly refrained from the statistical test interpretation, in stating that "[parallel analysis] must be viewed as a heuristic approach rather than a mathematically rigorous method resting on a solid conceptual base" (Crawford et al., 2010, p. 888) (Green, Levy, Thompson, Lu, & Lo, 2012, p. 360). In line with this, comparative simulation studies have shown an excellent performance of parallel analysis for $k=1$, and varying performance for $k > 1$ (Garrido, Abad, & Ponsoda, 2013; Jackson, 1993; Peres-Neto, Jackson, & Somers, 2005; Timmerman & Lorenzo-Seva, 2011).

Because Horn's parallel analysis is associated with PCA, rather than CFA, its use to indicate the number of common factors is inconsistent (Humphreys & Montanelli Jr, 1975; Ford, MacCallum, & Tait, 1986). Somewhat surprisingly, parallel analysis variants explicitly involving a common factor model perform clearly worse than (Crawford et al., 2010) or similar to (Green et al., 2012) Horn's parallel analysis in indicating the number of common factors. Many studies in this context include model error in their simulated data. Model error is represented by minor factors (Tucker, Koopman, & Linn, 1969) that account for a relatively small amount of variance in the common part of a variable. Horn's parallel analysis appears to indicate the number of major factors (Zwick & Velicer, 1982; Timmerman & Lorenzo-Seva, 2011), suggesting that Horn's parallel analysis is a reasonable heuristic for the number of major common factors. Note that the distinction between major and minor factors is arbitrary to some extent, and that also small factors may be important for interpretation in empirical practice.

Parallel analysis rests on resampling based distributions, because mathematical expressions for the eigenvalues distributions under the null model were not known yet in 1965. Using methods from random matrix theory, the asymptotic distribution for the largest eigenvalue of random covariance matrices was found by Johnstone (2001), who also provided an approximate statistical test for higher order components based on covariances. In the light of these recent developments it seems timely to review parallel analysis in the context of modern inferential statistics theory.

In this paper, we show that for the first eigenvalue parallel analysis is equivalent to a formal statistical test based on the Tracy-Widom distribution (Tracy & Widom, 1993, 1994). Further, we show that for the second and further eigenvalues parallel analysis is improper, as it is based on the joint distributions; in contrast, a Tracy-Widom based procedure is a proper test for the number of principal components based on covariances. For indicating the number of common factors, both methods are heuristic approaches at best. The theoretical relationships between the Tracy-Widom procedure and parallel analysis are illustrated using simulated data, complying with a principal component model. Further, the performance of both procedures for indicating the number of common factors is examined using simulated data complying with a common factor model, including major and minor factors. To illustrate the performance in empirical practice, both procedures are applied to data from two scale evaluation studies, and to data from a study

towards the intracultural structures underlying emotional experiences in different cultures.

The paper is organized as follows. First, we introduce the statistical background needed to arrive at a formal test for the significance of the largest eigenvalues of a sample covariance matrix. Then we introduce parallel analysis and establish its relationship to the Tracy-Widom procedure. Next we compare the two testing procedures on the basis of simulated data generated with different population covariance matrix structures, and using empirical data. We conclude by discussing the implications of our findings.

Background theory and theoretical setting

In this paper we focus explicitly on principal component analysis in the covariance setting. The major result illustrated here is Johnstone's Theorem, which concerns the asymptotic distribution of the largest eigenvalue of random covariance matrices. As this forms a reasonable approximation also for limited sample sizes (n) and number of variables (p), it is applicable as a statistical test for determining the number of principal components of empirical data. An analogous procedure for standardized data (*i.e.*, principal components based on correlations) is lacking so far. Though Johnstone's Theorem has been recently extended to the largest eigenvalue of random correlation matrices (Bao, Pan, & Zhou, 2012; Pillai, Yin, et al., 2012), this asymptotic result requires very large n and p to arrive at a reasonable approximation. An approximate solution has been proposed for the first largest eigenvalue, which appears applicable to smaller n and p (Saccenti, Smilde, Westerhuis, & Hendriks, 2011). However, no satisfying approach is available for testing the number of principal components in the case of correlation matrices.

Definitions

Given a data matrix \mathbf{X} containing n observations (rows) of p variables (columns), the sample covariance matrix \mathbf{C} is defined as

$$\mathbf{C} = \mathbf{X}^T \mathbf{X}. \quad (1)$$

If the entries of \mathbf{X} are identically and independently multivariate normally distributed, *i.e.*, $\mathbf{X} \sim \mathcal{N}(\mu, \Sigma)$, the sample covariance matrix \mathbf{C} is said to have a Wishart distribution with n degrees

of freedom and $p \times p$ population covariance matrix Σ , i.e., $\mathbf{C} \sim \mathcal{W}_p(n, \Sigma)$ (Wishart, 1928).

We distinguish the p population eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_p$ of the population covariance matrix Σ , and the p sample eigenvalues l_1, l_2, \dots, l_p of the sample covariance matrix \mathbf{C} , where it is assumed that the n observations are randomly drawn from the population.

Testing for equality of eigenvalues

The problem of determining the number of components can be translated into the question whether the associated population eigenvalues are equal to each other. That is, the test for the significance of the k -th eigenvalue is on the null hypothesis (see, for instance, Bartlett (1950))

$$H_{0k} : \lambda_k = \lambda_{k+1} = \dots = \lambda_p, \quad (2)$$

$k = 1, \dots, p$. The justification of the test, also known as Bartlett's test, is that the first $k - 1$ eigenvalues may indicate some substantial component of variation and that the last $p - k + 1$ eigenvalues are associated with components of equal variation and thus reflect "noise" (Jolliffe, 2005). A useful null model to consider in this context is in terms of the following diagonal population covariance matrix:

$$\Sigma_\tau = \text{diag}(\tau_1^2, \tau_2^2, \dots, \tau_{k-1}^2, \tau^2, \dots, \tau^2), \quad (3)$$

where $\tau_1^2 \geq \tau_2^2 \geq \dots \geq \tau_{k-1}^2 > \tau^2$. This model, referred as spiked covariance model, can be used in all cases with equality of the last $p - k + 1$ eigenvalues. The assumption of diagonality of Σ is without loss of generality, since every data matrix \mathbf{X} can be transformed via an orthonormal transformation into a matrix with diagonal covariance matrix, without changing the eigenvalues.

Now the question arises how to test the null hypothesis

$$H_0 : \Sigma = \Sigma_\tau. \quad (4)$$

The problem can be addressed by sequentially testing for the null hypothesis H_0 for $k = 1, 2, \dots$ until it is not rejected anymore. To test for $k = 1$, an exact test is available. For $k > 1$, an approximate extension of the latter exact test, under a population covariance model defined by

Equation (3), is available. The formal test and its extension were first proposed by Johnstone (2001) who derived the distribution of the first largest eigenvalue of random covariance matrices. We begin by introducing Johnstone's theorem on which the theoretical results presented in this paper do rest.

Theorem 1. (Johnstone) Let the entries of \mathbf{X} ($n \times p$) be identically and independently multivariate normally distributed (i.e., $\mathbf{X} \sim \mathcal{N}(0, \mathbf{I})$), and l_1 be the largest eigenvalue of the $p \times p$ sample covariance matrix $\mathbf{C} = \mathbf{X}^T \mathbf{X}$, with $n = n(p)$ an increasing function of p , implying that the number of samples should increase with the number of variables. Define the centering and scaling parameters (Johnstone, 2006) as

$$\begin{aligned}\eta_{np} &= (\sqrt{n-1/2} + \sqrt{p-1/2})^2 \\ \xi_{np} &= (\sqrt{n-1/2} + \sqrt{p-1/2}) \left(\frac{1}{\sqrt{n-1/2}} + \frac{1}{\sqrt{p-1/2}} \right)^{\frac{1}{3}}\end{aligned}\tag{5}$$

If

$$\lim_{n \rightarrow \infty} \frac{p}{n} \rightarrow \gamma \geq 1\tag{6}$$

then the statistic

$$L_1 = \frac{l_1 - \mu_{np}}{\sigma_{np}} \xrightarrow{\text{dist}} F_1(s, 1)\tag{7}$$

where $\mu_{np} = \eta_{np}$, $\sigma_{np} = \xi_{np}$ and $F_1(s, 1)$, is the Tracy-Widom distribution of order 1.

For a mathematical definition of the Tracy-Widom distribution, we refer the reader to Appendix 1. The theorem is stated for the case $n \geq p$, but it holds also for $n < p$: it is enough to exchange n and p in Equation (5). Figure 1 displays the probability density function (*pdf*) of the Tracy-Widom distribution of order 1 (panel A), and illustrates the functioning of Johnstone's theorem (panels B and C). This theorem holds true asymptotically, that is for data matrix dimensions going to infinity. However, numerical simulations have shown that the Tracy-Widom

limit holds even for the finite case, with n and p as small as 5 (Johnstone, 2001). Therefore, the Tracy-Widom distribution provides an excellent approximation of the largest eigenvalue distributions of a covariance matrix in the case of a finite sample matrix, which is of great interest in multivariate statistics and data analysis. Recently, Chiani (2012) provided semi-analytical expressions for the exact distribution of the largest eigenvalue of the covariance matrix for any given n and p , confirming the excellent agreement of the finite sample case with the asymptotic Tracy-Widom limit.

Johnstone's theorem provides the distribution of the L_1 statistics, and thus of l_1 , for the case with a population covariance matrix $\Sigma = \mathbf{I}$. From this, the L_1 -distribution can be derived for the more general population covariance matrix $\Sigma = \tau^2 \mathbf{I}$. This can be done by noting that $l_1(\tau^2 \mathbf{I}) = \tau^2 \cdot l_1(\mathbf{I})$, and thus the statistic L_1 can be computed using $\mu_{np} = \tau^2 \eta_{np}$ and $\sigma_{np} = \tau^2 \xi_{np}$.

To test for the first eigenvalue, one examines the null-hypothesis $H_0 : \Sigma = \tau^2 \mathbf{I}$, with τ^2 the population variances to test. (As will be clarified later, τ^2 can also be estimated on the basis of the sample data, but this results in an approximate, rather than an exact test.) For a given sample data set, one needs the statistics L_1 for the data at hand, and the Tracy-Widom percentile associated with the desired significance level α . Critical values can be obtained using software (*i.e.*, the RMstat package implemented in R (Johnstone, Ma, Perry, & Shahram, 2009); the RMTFredholmToolbox implemented in Matlab (Bornemann, 2009, 2010)), or from Table 1.

The formal testing procedure to assess the significance of the first component in principal component analysis can be summarized as follows:

1. Given a $n \times p$ data matrix \mathbf{X} , compute the $p \times p$ sample covariance matrix \mathbf{C} .
2. Compute the largest eigenvalue l_1 of \mathbf{C} .
3. Compute L_1 from l_1 , as specified by Johnstone's theorem in Equations (5) and (7), using $\mu_{np} = \tau^2 \eta_{np}$ and $\sigma_{np} = \tau^2 \xi_{np}$.
4. Compare L_1 with the appropriate Tracy-Widom percentile to test H_0 at significance level α .

Soshnikov (2002) has demonstrated that Johnstone's theorem also holds true for the case in which the data matrix entries have an arbitrary symmetric distribution with finite moments whose tails decay towards infinity at least as fast as those of a Gaussian distribution. This implies that

the constraint of normality can be relaxed, making Johnstone's theorem applicable in many situations encountered in empirical data analysis. Further generalizations and results concern rectangular matrices (Soshnikov, 2002), non-identity of the population covariance matrix (Baik, Ben Arous, & P      , 2005; Baik & Silverstein, 2006; Karoui, 2007) and different asymptotic behavior depending on the ratio n/p (and p/n) (Karoui, 2007).

We note that the definition of the sample covariance matrix \mathbf{C} in Equation (1) is common in random matrix theory. This contrasts with the standard definition

$$\mathbf{C}^* = \frac{1}{n-1}(\mathbf{X} - \mathbf{1}\mathbf{1}^T\mathbf{X})^T(\mathbf{X} - \mathbf{1}\mathbf{1}^T\mathbf{X}) \quad (8)$$

with $\mathbf{1}$ a $(n \times 1)$ vector of ones. This is not limiting, because it is proved that the asymptotic distributions of the largest eigenvalues of \mathbf{C}^* have the same distributional properties as \mathbf{C} , provided that $\mathbf{C} \sim \mathcal{W}_p(n, \mathbf{I})$ (Pan, 2012).

When considering the Tracy-Widom test for the first component, it is important to recognize that a limit of detection has been identified (Theorems 1.1, 1.2 and 1.3 in Baik & Silverstein, 2006). That is, the threshold on λ_1 is given by

$$\lambda_1 > 1 + \sqrt{\frac{p}{n}} \quad (9)$$

where n is the sample size and p is the number of variables. If the first population eigenvalue is below the threshold, the corresponding sample eigenvalues are Tracy-Widom distributed, and thus the first component is not distinguishable from noise (Baik et al., 2005). If it is above the threshold the sample eigenvalues will be unbounded due to a so-called phase transition (see for more details Appendix 2).

Testing for equality of higher order eigenvalues

Virtually in all situations of concern in multivariate analysis there is more than one eigenvalue standing out from the rest of the sample eigenvalues. The covariance model proposed in Equation (3) reasonably describes this condition and allows the definition of a statistical procedure to test for the 2^{nd} , 3^{rd} , etc. eigenvalue of a sample covariance matrix.

The core idea developed by Johnstone is to test against the null hypothesis under which the eigenvalues are consistent with those of a scaled Wishart distribution $\mathcal{W}_{p-k+1}(n, \tau^2 \mathbf{I})$, where the scale parameter τ^2 can be estimated from the data. This can be done using the relation

$$\hat{\tau}^2 = \frac{1}{n(p-k+1)} \sum_{q=k}^p l_q \quad (10)$$

To assess the null hypothesis $H_0 : \lambda_k = \lambda_{k+1} = \dots = \lambda_p$ at a given α level, the $\hat{x}_{100-\alpha}$ percentile of the approximated distribution for λ_k is calculated using the relation

$$\hat{x}_{100-\alpha} = \hat{\tau}^2 (\mu_{n,p-k+1} + x_{100-\alpha} \sigma_{n,p-k+1}) \quad (11)$$

where $x_{100-\alpha}$ is the $100 - \alpha$ percentile of the Tracy-Widom density distribution, $\mu_{n,p-k+1}$ and $\sigma_{n,p-k+1}$ are given by Equations (5) and $\hat{\tau}^2$ is the estimation of the scale parameter τ^2 . Hence, to assess the significance of the l_k eigenvalue at a given significance level α it is enough to confront it with the estimated percentile given by Equation (11). Johnstone (Proposition 1.2, 2001) has demonstrated that this approach is always conservative, at least for true Wishart matrices. Note that this approximate testing procedure could be applied to test for the first eigenvalue, implying that one estimates τ^2 for $k = 1$, on the basis of the sample data, rather than specifying it as in Equation (3). Matlab code for performing the TW test and parallel analysis is available at semantics.systemsbiology.nl.

Relationship between parallel analysis and the Tracy-Widom procedure

Parallel analysis (Horn, 1965) was proposed as an improvement of the Kaiser-Guttman-root-one criterion (Guttman, 1954) that, due to sampling fluctuations, suffers from an overestimation of the number of components. In particular, Horn proposed to use the mean of eigenvalues generated from independent normal variates as the threshold, rather than the value of 1, as in the Kaiser-Guttman-root-one criterion. Page 181 of his seminal paper (Horn, 1965) on parallel analysis reads (using Horn's original notation):

"Suppose that an investigator has obtained m measurements on N subjects. Call these measurements 'real data'. Now generate K matrices of random variables, each matrix of

order m by N ; intercorrelate the rows of these matrices; find the latent roots for the resulting K matrices; average (over K) the first root values. . . "

If one would replace the phrase *average (over K) the first root values...* with *calculate the $100 - \alpha$ percentile* (which is the modification proposed by Glorfeld (1995)) and use covariances instead of correlations (*intercorrelate*), the relationship between parallel analysis and the Tracy-Widom procedure becomes clear. That is, both test the null hypothesis for the first eigenvalue, using different approaches to arrive at the required sampling distribution. Summarizing, the parallel analysis procedure is as follows (using a notation consistent with the rest of this paper):

1. Given a $n \times p$ data matrix \mathbf{X} compute the $p \times p$ sample covariance matrix $\mathbf{C} = \mathbf{X}^T \mathbf{X}$ and the largest eigenvalue l_1 of \mathbf{C} .
2. Generate a random matrix \mathbf{X}^r of size $n \times p$ whose entries are *i.i.d* $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_\tau)$, compute its covariance matrix and compute its (first) largest eigenvalue l_1^r .
3. Repeat Step 2. $N \gg 1$ times and collect the N eigenvalues to build the reference null distribution of l_1^r .
4. Calculate the $100 - \alpha$ percentile from the reference distribution.
5. Compare l_1 with $100 - \alpha$ percentile to test H_0 at significance level α .

For the first eigenvalue, the Tracy-Widom test and parallel analysis test the same hypothesis: where the Tracy-Widom test is based on random matrix asymptotic theory, parallel analysis estimates the sampling distribution using resampling. The equivalence implies that parallel analysis is grounded on a solid statistical basis and should be considered a fully inferential approach.

It is interesting to note that parallel analysis has been found to be robust to departures from normality (Buja & Eyübgölu, 1992; Dinno, 2009; Timmerman & Lorenzo-Seva, 2011). Those results echo the universality theorems for the Tracy-Widom limit (Soshnikov, 2002) which disposed, under certain conditions, of the normality required by Johnstone's theorem.

As an alternative to drawing from $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_\tau)$ —to be denoted as parallel analysis-random in what follows—, a nonparametric approach has been proposed to obtain the eigenvalue sampling

distribution (Buja & Eyübgölu, 1992). We denote this approach as parallel analysis-permute in what follows, because a permutation of the raw data is used. For testing the significance of the first eigenvalue, PA-permute is similar to the approximate Tracy-Widom procedure for $k = 1$.

When parallel analysis is applied to correlation matrices, one uses either parallel analysis-random with $\Sigma_\tau = \mathbf{I}$, or parallel analysis-permute. The similarity in results of the two approaches in a simulation study (Timmerman & Lorenzo-Seva, 2011) can be explained by the fact that the associated population covariance matrices are equal.

Relationship between parallel analysis and the Tracy-Widom procedure for higher order components

Parallel analysis has been also used for the second eigenvalue and further (simply change l_1 to l_2 , and l_1^r, \dots to l_2^r, \dots , in Steps 2 to 5 above). In this case the statistical test interpretation is no longer straightforward, as indicated by various authors (Buja & Eyübgölu, 1992; Gorfeld, 1995; Crawford et al., 2010; Green et al., 2012). The empirical distributions for the $2^{nd}, 3^{rd}, \dots$ eigenvalue, generated under a null model using parallel analysis, correspond to the joint distribution of the $2^{nd}, 3^{rd}, \dots$ eigenvalues of random covariance matrices. Remarkably, these distributions are members of the Tracy-Widom distributions family, as shown by Shoshnikov's theorem (Shoshnikov, 2002):

Theorem 2. (Shoshnikov) Let the entries of the $n \times p$ data matrix \mathbf{X} be identically and independently distributed, i.e., $\mathbf{X} \sim \mathcal{D}(0, 1)$, with $n - p = O(n^{1/3})$ and $n/p \rightarrow \gamma \geq 1$, and distribution \mathcal{D} which is symmetric, with all finite moments, decaying at infinity at least as fast as a Gaussian distribution does.

Let l_k be the k -th largest eigenvalue of the $p \times p$ sample covariance matrix $\mathbf{C} = \mathbf{X}^T \mathbf{X}$. Define the centering and scaling parameters as in (5).

Then the statistic

$$L_k = \frac{l_k - \mu_{np}}{\sigma_{np}} \xrightarrow{dist} F_1(s, k) \quad (12)$$

where $F_1(s, k)$ is the Tracy-Widom distribution of order 1 for the k -th eigenvalue and μ_{np} and σ_{np} are defined as in Johnstone's Theorem.

This result provides the limiting joint distribution for the 2^{nd} , 3^{rd} , ... largest eigenvalues of random sample covariance matrices. As proven, this theorem holds for $0 \leq \gamma \leq \infty$ (Karoui, 2003).

In Figure 2, we provide for the 1^{st} to the 4^{th} largest eigenvalues examples of probability-probability plots for the Tracy-Widom joint distribution and the empirical distribution as obtained using parallel analysis. As can be seen, the agreement is excellent. This shows that for higher order eigenvalues ($k > 1$) the use of parallel analysis is equivalent to test the significance of the k^{th} eigenvalue using the joint distribution. We argue that this procedure is statistically debatable because to properly test the 2^{nd} , 3^{rd} , ... eigenvalues, one would need the distributions of the 2^{nd} , 3^{rd} , ... eigenvalues conditional upon the previous eigenvalues, rather than the joint distributions. This is so, because the subsequent eigenvalues are dependent upon each other (as the sum-of-eigenvalues equals the sum-of-squares). Thus, parallel analysis is unsuitable to test for higher order components, as one could already surmise based on the suboptimal performance for higher order components observed in the literature (Garrido et al., 2013; Jackson, 1993; Peres-Neto et al., 2005; Timmerman & Lorenzo-Seva, 2011). Unfortunately, analytical expressions for such distributions are not yet available.

Nevertheless, the approximated Tracy-Widom procedure proposed by Johnstone seems to be a valid and effective alternative, as shown by several authors (Patterson, Price, & Reich, 2006; Kritchman & Nadler, 2008) and by the results that we are going to present in the next sections.

Simulation study to compare the Tracy-Widom procedure and parallel analysis

To illustrate the theoretical relationships between the Tracy-Widom procedure and parallel analysis, we show their comparative performances on simulated data consistent with a principal component model. This is the covariance setting for which the Tracy-Widom procedure has been developed. Because parallel analysis is often used as a heuristic to indicate the number of major common factors, it might be of interest to see how the Tracy-Widom procedure performs in this context. Therefore we also considered simulated data that comply with a common factor model,

including major and minor factors. Note that for indicating the number of common factors both parallel analysis and the Tracy-Widom procedure must be considered as heuristic approaches.

Data generated under a principal component model

For the conditions consistent with the principal component model, we considered diagonal $p \times p$ population covariance matrices representing the null-model and the spiked model. In particular, for the null-model (model A) we considered a population covariance matrix Σ_A of the form

$$\Sigma_A = \text{diag}(\lambda_1, 1, 1, \dots, 1)_p \quad (13)$$

where we varied λ_1 as $1, 3, 5, \dots, 17, 20$, yielding 10 conditions, and took fixed $\lambda_k = 1$ for $2 \leq k \leq p$ (see Table 2, Σ -conditions 1 to 10). The condition with $\lambda_1 = 1$ is obviously equivalent to the null case defined by the Tracy-Widom distribution. In the conditions with $\lambda_1 > 1$, increasing values for λ_1 represent increasing departures from the null model.

For the spiked model (model B), we considered the population covariance matrix Σ_B as

$$\Sigma_B = \text{diag}(\lambda_1, \lambda_2, \lambda_3, 1, \dots, 1)_p \quad (14)$$

where we manipulated the magnitude of λ_1 , λ_2 and λ_3 to create different Σ -conditions. The actual values used are given in Table 2 (Σ -conditions 11 to 20).

For each population covariance matrix under consideration, we varied the data matrix dimensions at three levels (square ($n = p$), rectangular with $n < p$, rectangular with $n > p$), and the size of n, p at three levels (100, 300 and 1000), in a completely crossed design, using 1000 replicates per condition. It should be noted that such values of n and p yield a good approximation of the asymptotic properties outlined by Theorem 1, as has been shown (Johnstone, 2001, 2006).

Each simulated data matrix \mathbf{X} of size $n \times p$ is obtained by random sampling from a multivariate normal distribution $\mathcal{N}(0, \Sigma)$, taking $\Sigma = \Sigma_A$ or Σ_B , depending on the Σ -condition.

Data generated under a common factor model

For the population covariance matrices complying with a common factor model including both major and minor factors (model C), we used the same design as Timmerman and Lorenzo-Seva (2011). That is, we considered the population covariance matrix

$$\Sigma_C = w_{ma}\mathbf{\Lambda}_{ma}^T\mathbf{\Lambda}_{ma} + w_{mi}\mathbf{\Lambda}_{mi}^T\mathbf{\Lambda}_{mi} + w_{un}\mathbf{I}_p \quad (15)$$

where $\mathbf{\Lambda}_{ma}$ ($Q_{ma} \times p$) and $\mathbf{\Lambda}_{mi}$ ($Q_{mi} \times p$) are major and minor loading matrices, with Q_{ma} and Q_{mi} the number of major and minor factors, respectively; w_{ma} , w_{mi} and w_{un} are weights to manipulate σ_{ma}^2 , σ_{mi}^2 and σ_{un}^2 , the variances of the major part, minor part and unique part of the population covariance matrix Σ , respectively.

The major loading matrix $\mathbf{\Lambda}_{ma}$ was chosen fixed such that it relates each variable to a single major factor (with each row consisting of a single element equal to 1 and 0 otherwise). Each major factor is associated to 5 observed variables, hence the observed number of variables is $p = Q_{ma} \times 5$. The elements of the minor loading matrix $\mathbf{\Lambda}_{mi}$ are randomly drawn from a uniform distribution on the interval $[-1, 1]$. The number of minor factors was set at $Q_{mi} = Q_{ma} \times 3$.

We varied the number of major factors at 10 levels (taking $Q_{ma}=1,2,\dots,10$), the sample size at four levels (taking $n = 50, 100, 500, 1000$), and the relative size of the minor part at two levels (very small: $w_{mi} = \sqrt{0.1}$, $w_{ma} = \sqrt{0.75}$, $w_{un} = \sqrt{1 - 0.75 - 0.1}$ and small: $w_{mi} = \sqrt{0.35}$, $w_{ma} = \sqrt{0.5}$, $w_{un} = \sqrt{1 - 0.5 - 0.35}$). We used 100 replicates per condition. Each simulated data matrix \mathbf{X} of size $n \times p$ is obtained by random sampling from a multivariate normal distribution $\mathcal{N}(0, \Sigma_C)$.

Analyses of simulated data

Each simulated data matrix was analyzed with the stepwise Tracy-Widom testing procedure and with parallel analysis. For both, we took as the null distribution $\Sigma_\tau = \mathbf{I}$, in line with the simulation design.

Results and Discussion of the simulation study

Principal component model: Rejection rates

We start by assessing the empirical probability of rejecting H_0 considering a nominal α equal to 0.05, using parallel analysis and the Tracy-Widom procedure, for the eigenvalues at three types of positions, namely 1, 2 and 4; depending on the associated population eigenvalue, this probability indicates the actual α or the power of the test.

The rejection rates of parallel analysis and the Tracy-Widom procedure are summarized in graphical form in Figure 3 for the null-model condition (*i.e.*, $\Sigma = \Sigma_A$, Equation (13)) and Figure 4 for the spiked model condition (*i.e.*, $\Sigma = \Sigma_B$, Equation (14)). In both Figures, the successive rows refer to the conditions with the squared matrices 300×300 , rectangular matrices $n > p$, 300×100 , and rectangular matrices $n < p$, 100×300 . The comparative performances of both tests appeared to be highly similar for the other combinations of sizes (100, 300 and 1000). Therefore, those results are not summarized here; they can be obtained from the first author upon request. As can be seen in Figures 3 and 4, the comparative performances are highly similar across the three data matrix dimensions. Therefore, we only explicitly discuss the case for which $n = p = 300$ (*i.e.*, Figure 3, panels A and B; Figure 4, panels A, B and C).

In Figure 3, the rejection rates are displayed for the tests of l_1 (panel A) and l_2 (panel B) in the null-model condition (*i.e.*, $\Sigma = \Sigma_A$). When testing l_1 , it can be observed that the performances of the two methods appear to be highly similar, as expected. For λ_1 equal to one, the rejection rates for parallel analysis and the Tracy-Widom test were 0.061 and 0.059, respectively; thus the actual α of both tests is reasonably close to the nominal $\alpha = 0.05$. For both tests, with increasing λ_1 , the rejection rate increases, indicating an increasing power with increasing effect size, as expected. For parallel analysis, the sensitivity of the testing procedure depends also on how accurate is the empirical estimation of the percentiles of the distribution: we briefly comment on this in Appendix 3.

When testing l_2 (under $\lambda_2 = 1$), the rejection rates show striking differences between the Tracy-Widom test and parallel analysis. As can be seen in Figure 3, panel B, the Tracy-Widom test consistently shows actual α levels that are reasonably close to the nominal α of 0.05. In

contrast, the actual α levels of parallel analysis increase considerably with increasing levels of λ_1 , even up to about 0.5 in those conditions. Thus, the type I error rate of parallel analysis for the second eigenvalue increases with increasing size of the first population eigenvalue.

In Figure 4 the rejection rates are displayed for the tests of l_2 , l_3 , and l_4 (panels A, B, C, respectively) in the spiked model condition (i.e., $\Sigma = \Sigma_B$) with $n = p = 300$. Note that in all conditions considered, the tests for l_4 are under the null hypothesis, and for l_2 and l_3 under the alternative hypothesis. For all those tests, the rejection rates differ greatly between the Tracy-Widom procedure and parallel analysis.

As can be seen in Figure 4, panel C, the Tracy-Widom procedure for l_4 consistently shows actual α levels that are reasonably close to the nominal α of 0.05. In contrast, parallel analysis shows increasing type I error rates with increasing sizes of λ_1 (and thus of λ_2 and λ_3), even up to actual α levels of 1.

As can be seen in Figure 4, panels A and B, the power of both parallel analysis and the Tracy-Widom procedure increases with increasing population values (both for λ_2 and λ_3), as expected. Herewith, parallel analysis shows a consistently larger power than the Tracy-Widom procedure (except for those conditions where both tests have a power equal to 1). This seemingly favorable behavior of parallel analysis can be attributed to actual α levels that are way above the nominal ones.

The results in the null-model and spiked model conditions illustrate that parallel analysis to test for l_j , with $j > 1$, shows increasing type I error rates with increasing values of λ_k , $k < j$. This may be denoted as the pulling phenomenon, where with $\lambda_j = 1$, the sampling eigenvalues l_j increase with increasing values of λ_k , $k < j$. It is clear that parallel analysis is not able to address this situation: as previously illustrated, the use of parallel analysis to test high order eigenvalues is equivalent to the use of the joint distribution for the l_k largest eigenvalue that, of course, depends on the values of lower order eigenvalues. This means that the use of parallel analysis for testing the significance of second and higher order components is likely to result in the inclusion of noise in the PCA model.

Principal component model: Indicating the number of principal components

In Table 2 the averages (and standard errors) of the estimated number of components per testing procedure and Σ -condition are presented. The results clearly show that the two testing procedures perform about equally well when data are generated under the model with 0 components (*i.e.*, $\Sigma = \mathbf{I}$, condition 1). In all other conditions, with population dimensionality equal to or larger than 1, parallel analysis always overestimates the number of components. This observation is consistent with what has been shown in Figure 3. Moreover the variability observed in the estimated number of components for parallel analysis is extremely high, as indicated by the standard error (se), with coefficients of variation (se/mean) ≈ 1 or larger for all conditions. Thus we can conclude that in a situation in which there is no information shared between variables (*i.e.*, under a spiked model) or when a low degree of correlation is expected, parallel analysis is not an advisable testing procedure.

Common factor model: Indicating the number of major and minor common factors

The results of the comparison on simulated data under a common factor model, including both major and minor factors, are presented in Table 3. In the condition with a very small minor part (upper part of the Table), parallel analysis almost perfectly indicates the correct number of major factors across all conditions considered, except for the condition with a sample size as small as $n = 50$. The Tracy-Widom procedure shows similar behavior to parallel analysis for sample sizes up to $n = 100$. For the large sample sizes ($n = 500$ and $n = 1000$) the Tracy-Widom procedure indicates numbers higher than the true number of major factors. In the condition with a small minor part (lower part of Table 3), both parallel analysis and the Tracy-Widom procedure generally indicate numbers higher than the true number of major factors. Herewith, the Tracy-Widom procedure consistently indicates a larger number than parallel analysis. From both conditions, it appears that the Tracy-Widom procedure is more sensitive than parallel analysis to identify also small dimensions that stand out from noise.

Empirical examples to compare the Tracy-Widom procedure and parallel analysis

To illustrate the performance of the Tracy-Widom procedure and parallel analysis in empirical practice, we applied them to data from two scale evaluation studies and from a study on the structure of emotional experience. To each data set, we applied the Tracy-Widom procedure, parallel analysis-random with $\Sigma_r = \mathbf{I}$ and parallel analysis-permute. The latter two may yield different indications as the data were not standardized before analysis.

The Social Participation Questionnaire (SPQ) (Koster, Timmerman, Nakken, Pijl, & van Houten, 2009) was designed to assess four key themes of social participation of pupils with special needs in regular primary education. The SPQ consists of 24 items, where each item is associated with one key theme. Koster et al. (2009) hypothesized that the four presumed subscales would represent the four key themes and that the total SPQ scale would represent social participation in general. A re-analysis of the data collected on 590 pupils (see for details Koster et al. (2009)) revealed that the data could be represented adequately by a bifactor model (RMSEA = .069, CFI = .974) (Smits, Timmerman, & Meijer, 2012). This bifactor model had five common factors, with the general factor representing the total SPQ scale and four specific factors representing the four subscales. For this data, the Tracy-Widom procedure indicated in line with the bifactor model, namely five, whereas the parallel analyses indicated lower figures, namely three with parallel analysis-random, and two with parallel analysis-permute.

The Narcissistic Personality Inventory (NPI) (Raskin & Hall, 1979) is a commonly used instrument to measure narcissism as a personality trait in normal populations. The factorial structure has been debated, with the number of factors varying from two, three, four, or seven factors (Barelds & Dijkstra, 2010). For the Dutch adaptation of the 40-item NPI (Raskin & Terry, 1988), Barelds and Dijkstra (2010) examined the structure using data from a community sample ($n = 460$). Loading matrices from PCA's with four and seven components were procrustes rotated towards the hypothesized structure, but showed relatively low congruences (ranging from .52 to .89). Based on the interpretability of (obliquely rotated) loadings, Barelds and Dijkstra (2010) favored a two-component solution. For this data, the Tracy-Widom procedure and the parallel analyses indicated larger figures, namely seven with the Tracy-Widom procedure, six with parallel

analysis-random and four with parallel analysis-permute.

In a study towards similarities in intracultural and intercultural dimensions in recalled emotional experience, the structure of the self-reported frequency of the experience of different emotions within and across nations was examined (Kuppens, Ceulemans, Timmerman, Diener, & Kim-Prieto, 2006). The data pertain to ratings on a 9-point scale, on 14 emotions (including e.g., pleasant, happy, anger, guilt, shame), which are available of 9300 participants stemming from 48 different nations. We applied the Tracy-Widom procedure and the parallel analyses to the data per country. This implies that we examine the number of dimensions for each country separately. This contrasts to Kuppens et al. (2006), who focused on the dimensions that are common across all countries, using a simultaneous component analysis. Using a scree-plot, they identified two components. Differences between the two approaches may occur if some dimensions are small and/or rather specific for one, or a few countries: These dimensions will be kept hidden in the simultaneous component analysis (as in Kuppens et al. (2006)), and may be identified using an analysis per country.

As can be seen in Table 4, the number of components indicated differ substantially between methods. Parallel analysis-random consistently indicates very large numbers of components (i.e., ranging from 9 to 14). This suggests that the population covariance matrix of $\Sigma_{\tau} = \mathbf{I}$ is rather unsuitable in this context. This is not very suprising, given that the observed variances across inhabitants per emotion and country range from 1.17 to 9.45 (median 3.89). Parallel analysis-permute and the Tracy-Widom procedure indicate numbers of components ranging from zero to four. Herewith, the Tracy-Widom procedure indicates either the same or a higher number of components, with a maximal difference of two.

Discussion

Parallel analysis is widely used tool for assessing the dimensionality of empirical data. Since its introduction it has gained ample popularity, especially in the field of social and behavioral sciences (Fabrigar et al., 1999; Hayton et al., 2004; Thompson, 2004). In this paper we have shown that when used to assess the significance of the first principal component based on covariances, parallel analysis is just an empirical realization of the Tracy-Widom test. Our results show indeed

that the two procedures are equivalent and lead to the same experimental results.

Based on theoretical considerations and numerical simulations we also showed that the use of parallel analysis to test the significance of higher order components leads to an overestimation of the number of components, while a testing procedure based on the Tracy-Widom distribution is able to correctly address the problem of the rejection of non-significant components associated to noise. This is so because the Tracy-Widom procedure for higher order components is based on conditional distributions rather than the joint distribution. Green *et al.* (2012) proposed parallel analysis variants for testing the $k^{th} + 1$ common factor or principal component, while incorporating a model for k^{th} common factor or principal component. Those variants can be seen as a resampling based estimation of the conditional distribution. The variant for principal components is a direct counterpart of the Tracy-Widom procedure.

Parallel analysis is often applied to obtain an indication of the number of major common factors. Our simulations under the common factor model suggest that parallel analysis yield variable results, with typically correct indications of the number of major factors in conditions with a very small minor part, and numbers higher than the number of major factors with a small minor part. The Tracy-Widom procedure consistently indicates a larger number than the true number of major factors. We comment here that the Tracy-Widom procedure is not able to distinguish between major and minor factors because it tests for relevant dimensions (which include both major and minor factors) other than those associated to noise. As a matter of fact, also parallel analysis is not able to operate such a distinction: such a procedure is equivalent to the use of the joint distribution of high order eigenvalues for testing their significance and the distinction between major and minor factors is by no means encoded in the joint distribution.

In the light of this, we argue that this apparent capability of parallel analysis of correctly indicating the number of major factors, and discarding the minor ones, is only a consequence of the reduced power of parallel analysis. Stated otherwise, minor factors are discarded because the parallel analysis procedure does not have enough power to detect them, not because parallel analysis would focus on major factors. This can be seen in Figure 5, where the population eigenvalues and sample eigenvalues of a simulated data set for the condition with a very small minor part are presented, for the conditions with the number of major factors $Q_{ma} = 1, \dots, 9$. The

major factors clearly stand out, but also minor factors separate to some extent from background noise. Therefore, in the context of a common factor model, both the Tracy-Widom procedure and parallel analysis should be viewed as heuristic procedures at best.

When the aim is to identify the number of major factors or principal components, it seems obvious to use a criterion explicitly designed for this goal. The CHull (Ceulemans & Kiers, 2006) for model selection is such a criterion. It detects a model with an optimal balance between a (large) model fit and (low) number of parameters. To indicate the number of major principal components, the CHull criterion can be directly applied (Wilderjans, Ceulemans, & Meers, 2013). To indicate the number of major common factors, one can apply the Hull method, which is based on the CHull criterion. In a simulation study the Hull method appeared to outperform parallel analysis in indicating the number of major common factors (Lorenzo-Seva, Timmerman, & Kiers, 2011).

Appendices

Appendix 1: The Tracy-Widom distribution

The Tracy-Widom distribution (Tracy & Widom, 1993, 1994, 1996) is defined as

$$F_1(s, 1) = \exp \left(-\frac{1}{2} \int_s^\infty q(t) + (t-s)q^2(t)dt \right) \quad (16)$$

The function $q(x)$ is the unique Hastings-McLeod solution (Hastings & McLeod, 1980) of the non-linear Painlevé differential equation

$$\frac{d^2 q}{dt^2} = tq(t) + 2q^3(t)$$

satisfying the boundary condition

$$q(t) \approx Ai(t) \text{ as } x \rightarrow -\infty$$

where $Ai(t)$ is the Airy function (Airy, 1838) and

$$q(x) = \sqrt{-\frac{x}{2}} \left(1 + \frac{1}{8x^3} + O\left(\frac{1}{x^6}\right) \right) \text{ as } x \rightarrow -\infty$$

This distribution was found to be the limiting law for the largest eigenvalue of Gaussian symmetric $n \times n$ matrices (the so-called GOE, Gaussian Orthogonal Ensemble). Johnstone's theorem showed that the same limiting distribution holds for the covariance matrices of rectangular data matrices $n \times p$ when both n and p are large.

Appendix 2: The Baik-Ben Arous-Péché phase transition

Baik et al. (2005) provided the asymptotics of the distribution of the largest eigenvalue(s) of a sample covariance matrix under the spike population model. These results were proved for complex data, but there is strong evidence that they hold true also for real data. For this, the results in Baik et al. (2005) can be stated in form of a conjecture, the so called Baik-Ben Arous-Péché (BBP conjecture).

Conjecture 1. Let λ_1 be the leading eigenvalue of the population covariance matrix with $\lambda_k = 1$ for $2 \leq k \leq p$ and let l_1 be the largest sample eigenvalue. In the asymptotic regime $(n, p) \rightarrow \infty$ with finite limit ratio p/n it holds that:

1. If

$$\lambda_1 < 1 + \sqrt{\frac{p}{n}} \quad (17)$$

then l_1 when properly normalized to L_1 , will have the same distribution as when $\lambda_1 = 1$, that is it will be Tracy-Widom distributed.

2. If

$$\lambda_1 \geq 1 + \sqrt{\frac{p}{n}} \quad (18)$$

then L_1 will be almost surely unbounded.

Statement (2) was proved for real data in Baik and Silverstein (2006) and Paul (2007) showed that in the real case, L_1 will exhibit Gaussian fluctuations.

The behavior of L_1 will be different depending on the size of λ_1 , hence the phase-transition denomination. It is clear that as $(n, p) \rightarrow \infty$ the phase transition become arbitrary sharp. Stated otherwise, if λ_1 is below the BBP limit there will be little chance to detect structure in the data, as the eigenvalues are distributed according the Tracy-Widom distribution, that is, as noise eigenvalues. On the contrary, if λ_1 is above the BBP limit detection of structure will be easier (Patterson et al., 2006). This phenomenon has been recently used to explain some problems arising in eigenanalysis when applied to population genetic studies (Patterson et al., 2006) and economics (Harding, 2008). Similar results hold for higher order eigenvalues: it is enough to replace l_1 and λ_1 with l_k and λ_k in Equations (17) and (18). For more details, see Karoui (2007), Johnstone (2006), Paul (2007), Paul & Aue (2014), Tracy & Widom (2009).

An interesting aspect of the BBP phase transition is that it provides one of the few examples of power analysis in the multivariate setting. Given a problem and fixed the population size p , if λ_1 would have been known, then Equation (18) would give a direct estimate of the sample size needed to be able to detect the presence of structure in the data. From Equation (18) also descends that, increasing sample size, rather than variable number, is advantageous for detecting

structure above the BBP threshold, but if λ_1 lays below there is no gain in increasing the sample size (Patterson et al., 2006).

Appendix 3: A note on percentiles estimation

Here we give a brief look at the quality of the estimates of the percentiles of the null distribution, as obtained using parallel analysis. Consistently with the rest of the paper we consider the percentiles of the Tracy-Widom distribution as the target against which to compare the estimations obtained using parallel analysis. For the Tracy-Widom distribution, the percentiles can be calculated with almost arbitrary precision using computational toolboxes (see Bornemann, 2009, 2010). In contrast, obtaining reliable estimations of the percentiles of a population density function from an empirical distribution is a known problem even when the parameter of the population density (*i.e.*, mean and variance) are known (Efron & Tibshirani, 1993; Efron, 1994; Rice & Church, 1996; DiCiccio & Efron, 1996). Using simulations we found that the error on the percentiles estimation decreases, as expected, with the number N of realizations used to build the empirical distribution. This is also indicated by the formula:

$$\sigma_{100-\alpha,k} = \frac{\sigma_{TW}}{y_{100-\alpha}} \sqrt{\frac{P(1-P)}{N}} \quad (19)$$

where $y_{100-\alpha} = TW^{-1}(x_{100-\alpha})$, σ_{TW} is the standard deviation of $TW_1(s)$, and P is the proportion $100 - \alpha$ (see *e.g.*, Kendall & Yule, 1950; Deming, 1966; Rice & Church, 1996)).

Figure 6 shows the results of a simulation, where the 95% and 99% percentiles of the Tracy-Widom distribution were estimated using different numbers of realizations N , using 10 replicates. As anticipated, the error decreases with N . Another aspect to be considered is the sampling variability: given a fixed N , different realizations may lead to (slightly) different estimations of the percentiles and this variability can be really large if the number of realizations is limited. This is clearly shown in Figure 6: estimations tend to be really unstable for small numbers of realizations ($N < 10^3$) (panel A) and even for size of $N = 10^5$, we can still observe instability of the estimations (panel D).

With $N = 10^5$ the relative error is slightly below 1%. In the literature some authors have recommended $N = 300$ in (Ledesma & Valero-Mora, 2007) or $N = 2500$ (Buja & Eyübgölu, 1992)

being the latter a number that can still lead to a relative error in the range of 10%. Theoretical estimates of the error on the estimation can be obtained using the Yule-Kendall formula. This is outside the scope of this paper, but preliminary simulations seems to indicate that at least $N > 10^7$ are needed to obtain a relative precision lower then 0.1%.

Notation

Summary of the notation used in the paper:

\mathbf{X}	matrices (bold font, uppercase)
x_{ij}	element of \mathbf{X} in the i -th row, j -th column
Σ	population covariance matrix
\mathbf{C}	sample covariance matrix
λ_k	k -th eigenvalue of the population covariance matrix
l_k	k -th eigenvalue of the sample covariance matrix
L_k	Tracy-Widom statistic for the l_k
s	argument of the Tracy-Widom <i>cdf</i> and <i>pdf</i>

Acknowledgments

The authors thank Dick Barelds, Marloes Koster and Sip Jan Pijl for sharing their data. This work was partly supported by the European Commission-funded FP7 project INFECT (Contract No.305340).

References

- Airy, G. (1838). On the intensity of light in the neighbourhood of a caustic. *Transactions of the Cambridge Philosophical Society*(6), 379–402.
- Baik, J., Ben Arous, G., & Péché, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability*, 1643–1697.
- Baik, J., & Silverstein, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6), 1382–1408.
- Bao, Z., Pan, G., & Zhou, W. (2012). Tracy-widom law for the extreme eigenvalues of sample correlation matrices. *Electron. J. Probab*, 17(88), 1–32.
- Barelds, D. P., & Dijkstra, P. (2010). Narcissistic personality inventory: structure of the adapted dutch version. *Scandinavian journal of psychology*, 51(2), 132–138.
- Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of statistical psychology*, 3(2), 77–85.
- Bornemann, F. (2009). On the numerical evaluation of distributions in random matrix theory: a review. *arXiv preprint arXiv:0904.1581*.
- Bornemann, F. (2010). On the numerical evaluation of fredholm determinants. *Mathematics of Computation*, 79(270), 871–915.
- Buja, A., & Eyübgölu, N. (1992). Remarks on parallel analysis. *Multivariate behavioral research*, 27(4), 509–540.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate behavioral research*, 1(2), 245–276.
- Ceulemans, E., & Kiers, H. A. (2006). Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method. *British journal of mathematical and statistical psychology*, 59(1), 133–150.
- Chiani, M. (2012). Distribution of the largest eigenvalue for real wishart and gaussian random matrices and a simple approximation for the tracy-widom distribution. *arXiv preprint arXiv:1209.3394*.
- Crawford, A. V., Green, S. B., Levy, R., Lo, W.-J., Scott, L., Svetina, D., & Thompson, M. S. (2010). Evaluation of parallel analysis methods for determining the number of factors. *Educational and Psychological Measurement*.
- Deming, W. E. (1966). *Some theory of sampling*. Courier Dover Publications.
- DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical science*, 189–212.
- Dinno, A. (2009). Exploring the sensitivity of horn’s parallel analysis to the distributional form of random data. *Multivariate behavioral research*, 44(3), 362–388.
- Efron, B. (1994). Missing data, imputation, and the bootstrap. *Journal of the American Statistical Association*, 89(426), 463–475.
- Efron, B., & Tibshirani, R. J. (1993). The bootstrap estimate of standard error. In *An introduction to the bootstrap* (pp. 45–59). Springer.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological methods*, 4(3), 272.
- Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel psychology*, 39(2), 291–314.

- Garrido, L. E., Abad, F. J., & Ponsoda, V. (2013). A new look at horn's parallel analysis with ordinal variables. *Psychological methods*, 18(4), 454.
- Glorfeld, L. W. (1995). An improvement on horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and psychological measurement*, 55(3), 377–393.
- Green, S. B., Levy, R., Thompson, M. S., Lu, M., & Lo, W.-J. (2012). A proposed solution to the problem with using completely random data to assess the number of factors with parallel analysis. *Educational and Psychological Measurement*, 72(3), 357–374. Retrieved from <http://epm.sagepub.com/content/72/3/357.abstract> doi: 10.1177/0013164411422252
- Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, 19(2), 149–161.
- Harding, M. C. (2008). Explaining the single factor bias of arbitrage pricing models in finite samples. *Economics Letters*, 99(1), 85–88.
- Hastings, S., & McLeod, J. (1980). A boundary value problem associated with the second Painleve transcendent and the Korteweg-de Vries equation. *Archive for Rational Mechanics and Analysis*, 73(1), 31–51.
- Hattie, J. (1985). Methodology review: assessing unidimensionality of tests and Itenls. *Applied Psychological Measurement*, 9(2), 139–164.
- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational research methods*, 7(2), 191–205.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179–185.
- Humphreys, L. G., & Montanelli Jr, R. G. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research*, 10(2), 193–205.
- Jackson, D. A. (1993). Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology*, 2204–2214.
- Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Annals of statistics*, 295–327.
- Johnstone, I. M. (2006). High dimensional statistical inference and random matrices. *arXiv preprint math/0611589*.
- Johnstone, I. M., Ma, Z., Perry, P. O., & Shahram, M. (2009). Rmtstat: Distributions, statistics and tests derived from random matrix theory [Computer software manual]. (R package version 0.2)
- Jolliffe, I. (2005). *Principal component analysis*. Wiley Online Library.
- Karoui, N. E. (2003). On the largest eigenvalue of wishart matrices with identity covariance when n , p and p/n tend to infinity. *arXiv preprint math/0309355*.
- Karoui, N. E. (2007). Tracy-widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *The Annals of Probability*, 663–714.
- Kendall, M. G., & Yule, G. U. (1950). *An introduction to the theory of statistics*. Charles Griffin & Company.
- Koster, M., Timmerman, M. E., Nakken, H., Pijl, S. J., & van Houten, E. J. (2009). Evaluating social participation of pupils with special needs in regular primary schools. *European*

- Journal of Psychological Assessment*, 25(4), 213–222.
- Kritchman, S., & Nadler, B. (2008). Determining the number of components in a factor model from limited noisy data. *Chemometrics and Intelligent Laboratory Systems*, 94(1), 19–32.
- Kuppens, P., Ceulemans, E., Timmerman, M. E., Diener, E., & Kim-Prieto, C. (2006). Universal intracultural and intercultural dimensions of the recalled frequency of emotional experience. *Journal of Cross-Cultural Psychology*, 37(5), 491–515.
- Ledesma, R. D., & Valero-Mora, P. (2007). Determining the number of factors to retain in efa: an easy-to-use computer program for carrying out parallel analysis. *Practical Assessment, Research & Evaluation*, 12(2), 1–11.
- Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. (2011). The hull method for selecting the number of common factors. *Multivariate Behavioral Research*, 46(2), 340–364.
- Pan, G. (2012). Comparison between two types of large sample covariance matrices. *Ann. Institut Henri Poincaré*.
- Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS genetics*, 2(12), e190.
- Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, 17(4), 1617.
- Paul, D., & Aue, A. (2014). Random matrix theory in statistics: A review. *Journal of Statistical Planning and Inference*, 150(0), 1–29.
- Peres-Neto, P. R., Jackson, D. A., & Somers, K. M. (2005). How many principal components? stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics & Data Analysis*, 49(4), 974–997.
- Pillai, N. S., Yin, J., et al. (2012). Edge universality of correlation matrices. *The Annals of Statistics*, 40(3), 1737–1763.
- Raskin, R., & Hall, C. (1979). A narcissistic personality inventory. *Psychological reports*, 45(2), 590–590.
- Raskin, R., & Terry, H. (1988). A principal-components analysis of the narcissistic personality inventory and further evidence of its construct validity. *Journal of personality and social psychology*, 54(5), 890.
- Rice, S., & Church, M. (1996). Sampling surficial fluvial gravels: the precision of size distribution percentile estimates. *Journal of Sedimentary Research*, 66(3).
- Saccenti, E., Smilde, A. K., Westerhuis, J. A., & Hendriks, M. M. (2011). Tracy–widom statistic for the largest eigenvalue of autoscaled real matrices. *Journal of Chemometrics*, 25(12), 644–652.
- Smits, I. A., Timmerman, M. E., & Meijer, R. R. (2012). Exploratory mokken scale analysis as a dimensionality assessment tool why scalability does not imply unidimensionality. *Applied Psychological Measurement*, 36(6), 516–539.
- Soshnikov, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *Journal of Statistical Physics*, 108(5), 1033–1056.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. American Psychological Association.
- Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, 16(2), 209.

- Tracy, C. A., & Widom, H. (1993). Level-spacing distributions and the airy kernel. *Physics Letters B*, 305(1), 115–118.
- Tracy, C. A., & Widom, H. (1994). Level-spacing distributions and the airy kernel. *Communications in Mathematical Physics*, 159(1), 151–174.
- Tracy, C. A., & Widom, H. (1996). On orthogonal and symplectic matrix ensembles. *Communications in Mathematical Physics*, 177(3), 727–754.
- Tracy, C. A., & Widom, H. (2009). The distributions of random matrix theory and their applications. In *New trends in mathematical physics* (pp. 753–765). Springer.
- Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, 34(4), 421–459.
- Wilderjans, T. F., Ceulemans, E., & Meers, K. (2013). Chull: A generic convex-hull-based model selection method. *Behavior research methods*, 45(1), 1–15.
- Wishart, J. (1928). The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 32–52.
- Zwick, W. R., & Velicer, W. F. (1982). Factors influencing four rules for determining the number of components to retain. *Multivariate behavioral research*, 17(2), 253–269.

TABLES

<i>P</i> -value	TW-percentile	<i>P</i> -value	TW- percentile	<i>P</i> -value	TW-percentile	<i>P</i> -value	TW-percentile
0.0001	4.3508	0.0130	1.8658	0.0255	1.4408	0.0380	1.1726
0.0010	3.2713	0.0135	1.8428	0.0260	1.4281	0.0385	1.1636
0.0015	3.0659	0.0140	1.8205	0.0265	1.4155	0.0390	1.1546
0.0020	2.9170	0.0145	1.7990	0.0270	1.4032	0.0395	1.1458
0.0025	2.7994	0.0150	1.7781	0.0275	1.3911	0.0400	1.1371
0.0030	2.7020	0.0155	1.7578	0.0280	1.3792	0.0405	1.1284
0.0035	2.6187	0.0160	1.7381	0.0285	1.3674	0.0410	1.1199
0.0040	2.5458	0.0165	1.7189	0.0290	1.3558	0.0415	1.1114
0.0045	2.4808	0.0170	1.7003	0.0295	1.3444	0.0420	1.1030
0.0050	2.4222	0.0175	1.6822	0.0300	1.3332	0.0425	1.0947
0.0055	2.3687	0.0180	1.6644	0.0305	1.3221	0.0430	1.0865
0.0060	2.3195	0.0185	1.6472	0.0310	1.3112	0.0435	1.0784
0.0065	2.2740	0.0190	1.6303	0.0315	1.3005	0.0440	1.0703
0.0070	2.2315	0.0195	1.6138	0.0320	1.2898	0.0445	1.0623
0.0075	2.1918	0.0200	1.5977	0.0325	1.2794	0.0450	1.0545
0.0080	2.1544	0.0205	1.5820	0.0330	1.2691	0.0455	1.0466
0.0085	2.1190	0.0210	1.5666	0.0335	1.2588	0.0460	1.0389
0.0090	2.0856	0.0215	1.5515	0.0340	1.2488	0.0465	1.0312
0.0095	2.0537	0.0220	1.5367	0.0345	1.2389	0.0470	1.0236
0.0100	2.0234	0.0225	1.5222	0.0350	1.2290	0.0475	1.0160
0.0105	1.9943	0.0230	1.5080	0.0355	1.2194	0.0480	1.0086
0.0110	1.9666	0.0235	1.4941	0.0360	1.2098	0.0485	1.0011
0.0115	1.9399	0.0240	1.4804	0.0365	1.2003	0.0490	0.9938
0.0120	1.9143	0.0245	1.4669	0.0370	1.1910	0.0495	0.9865
0.0125	1.8896	0.0250	1.4538	0.0375	1.1817	0.0500	0.9793

Table 1: Percentiles for the Tracy-Widom distribution calculated by inverse interpolation of the cumulative distribution function.

Σ -condition	$\lambda_1, \lambda_2, \lambda_3, \lambda_4$	Number of estimated components							
		100×10		100×18		100×30		100×50	
		PA	TW	PA	TW	PA	TW	PA	TW
1	1, 1, 1, 1	0.1 (0.5)	0.0 (0.2)	0.1 (0.2)	0.0 (0.1)	0.2 (1.2)	0.0 (0.2)	0.1 (0.5)	0.0 (0.2)
2	3, 1, 1, 1	2.1 (1.5)	1.0 (0.0)	2.0 (1.5)	1.0 (0.1)	1.8 (1.8)	1.0 (0.0)	1.6 (1.4)	1.0 (0.2)
3	5, 1, 1, 1	1.7 (1.3)	1.0 (0.0)	1.8 (1.3)	1.0 (0.0)	1.8 (1.3)	1.0 (0.1)	2.0 (1.8)	1.0 (0.1)
4	7, 1, 1, 1	1.9 (1.4)	1.0 (0.0)	1.8 (1.5)	1.0 (0.0)	2.4 (1.9)	1.0 (0.2)	2.5 (2.2)	1.0 (0.2)
5	9, 1, 1, 1	1.8 (1.2)	1.0 (0.0)	2.4 (2.3)	1.0 (0.0)	2.3 (2.6)	1.0 (0.2)	2.3 (2.1)	1.0 (0.1)
6	11, 1, 1, 1	2.0 (1.9)	1.0 (0.0)	1.9 (1.6)	1.0 (0.0)	2.2 (2.0)	1.0 (0.1)	1.9 (2.0)	1.0 (0.1)
7	13, 1, 1, 1	1.8 (1.1)	1.0 (0.0)	2.1 (1.8)	1.0 (0.0)	2.5 (2.8)	1.0 (0.1)	1.7 (1.2)	1.0 (0.1)
8	15, 1, 1, 1	1.7 (1.1)	1.0 (0.0)	1.9 (1.3)	1.0 (0.0)	2.0 (1.6)	1.0 (0.0)	2.0 (1.3)	1.0 (0.2)
9	17, 1, 1, 1	1.7 (1.4)	1.0 (0.0)	1.9 (1.4)	1.0 (0.0)	2.2 (2.2)	1.0 (0.0)	2.0 (1.5)	1.0 (0.1)
10	20, 1, 1, 1	2.1 (1.8)	1.0 (0.0)	2.1 (1.9)	1.0 (0.0)	2.1 (2.0)	1.0 (0.0)	1.9 (1.4)	1.0 (0.0)
11	3, 2, 1, 1	4.0 (2.1)	1.4 (0.5)	3.9 (2.4)	1.5 (0.5)	4.0 (2.6)	1.4 (0.5)	4.2 (4.4)	1.1 (0.4)
12	5, 4, 1, 1	3.8 (1.8)	2.0 (0.0)	4.8 (3.1)	2.0 (0.0)	4.9 (2.6)	2.0 (0.0)	6.0 (4.2)	2.0 (0.1)
13	7, 6, 1, 1	4.3 (2.0)	2.0 (0.0)	4.5 (2.4)	2.0 (0.0)	5.0 (2.8)	2.0 (0.0)	5.4 (3.6)	2.0 (0.1)
14	9, 8, 1, 1	4.8 (2.3)	2.0 (0.0)	5.0 (3.2)	2.0 (0.0)	5.4 (3.4)	2.0 (0.0)	6.4 (4.3)	2.0 (0.1)
15	11, 10, 1, 1	4.7 (2.2)	2.0 (0.0)	5.5 (3.5)	2.0 (0.0)	5.5 (4.2)	2.0 (0.0)	6.7 (5.1)	2.0 (0.1)
16	5, 4, 3, 1	6.8 (2.2)	2.7 (0.5)	8.2 (3.8)	2.9 (0.3)	10.0 (5.5)	3.0 (0.2)	11.2 (5.9)	2.9 (0.3)
17	7, 6, 5, 1	6.6 (2.0)	3.0 (0.0)	7.7 (3.5)	3.0 (0.0)	11.0 (5.5)	3.0 (0.0)	11.7 (6.4)	3.0 (0.0)
18	9, 8, 7, 1	6.9 (2.2)	3.0 (0.0)	8.6 (4.2)	3.0 (0.0)	10.1 (5.3)	3.0 (0.0)	11.4 (6.3)	3.0 (0.1)
19	11, 10, 9, 1	6.9 (2.1)	3.0 (0.0)	9.9 (4.3)	3.0 (0.0)	9.6 (5.1)	3.0 (0.0)	12.9 (7.4)	3.0 (0.0)
20	13, 12, 11, 1	6.8 (2.1)	3.0 (0.0)	8.6 (3.5)	3.0 (0.0)	10.9 (5.2)	3.0 (0.0)	13.1 (6.7)	3.0 (0.0)

Table 2: Average number (and standard error) of estimated of principal components according to parallel analysis (PA) and the Tracy-Widom procedure (TW) (over 1000 data sets), in 20 types, with 3 data sizes. Data sets 1 to 10 are generated under a population covariance matrix $\Sigma = \Sigma_A$ (see Equation (13)). Data sets 11 to 20 are generated under a population covariance matrix $\Sigma = \Sigma_A$ (see Equation (14)). Values for $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are presented in column 2; $\lambda_k = 1$ for $k > 3$. Tests are at a nominal $\alpha = 0.05$ significance level.

Minor factors: very small		Estimated number of factors					
		$50 \times p$		$100 \times p$		$500 \times p$	
Q_{ma}	p	PA	TW	PA	TW	PA	TW
1	5	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)	1.1 (0.3)	1.0 (0.0)	1.7 (0.5)
2	10	2.0 (0.0)	2.0 (0.0)	2.0 (0.0)	2.2 (0.4)	2.0 (0.0)	2.9 (0.3)
3	15	3.0 (0.0)	3.0 (0.0)	3.0 (0.0)	3.1 (0.3)	3.0 (0.0)	4.0 (0.0)
4	20	4.0 (0.0)	4.1 (0.3)	4.0 (0.0)	4.0 (0.0)	4.0 (0.0)	5.1 (0.3)
5	25	4.8 (0.4)	5.1 (0.3)	5.0 (0.0)	5.1 (0.3)	5.0 (0.0)	6.1 (0.6)
6	30	5.7 (0.5)	6.0 (0.0)	6.0 (0.0)	6.2 (0.4)	6.0 (0.0)	7.5 (0.7)
7	35	6.5 (0.7)	7.0 (0.0)	7.0 (0.0)	7.0 (0.0)	7.0 (0.0)	8.8 (0.6)
8	40	6.9 (0.3)	8.0 (0.0)	8.0 (0.0)	8.0 (0.0)	8.0 (0.0)	9.4 (0.7)
9	45	7.4 (0.7)	9.0 (0.0)	8.9 (0.3)	9.0 (0.0)	9.0 (0.0)	10.4 (0.5)
10	50	8.6 (0.7)	10.1 (0.3)	10.0 (0.0)	10.0 (0.0)	10.0 (0.0)	11.5 (0.5)
Minor factors: small							
Q_{ma}	p	PA	TW	PA	TW	PA	TW
1	5	1.0 (0.1)	1.2 (0.4)	1.1 (0.3)	1.6 (0.5)	1.3 (0.4)	2.0 (0.2)
2	10	2.1 (0.4)	2.5 (0.5)	2.3 (0.4)	2.9 (0.4)	2.6 (0.5)	3.6 (0.5)
3	15	3.1 (0.5)	3.5 (0.5)	3.4 (0.6)	4.1 (0.5)	3.9 (0.5)	5.1 (0.3)
4	20	3.9 (0.6)	4.5 (0.6)	4.4 (0.5)	5.2 (0.5)	5.1 (0.6)	6.6 (0.5)
5	25	4.8 (0.7)	5.4 (0.7)	5.4 (0.6)	6.2 (0.6)	6.3 (0.6)	8.1 (0.5)
6	30	5.5 (0.8)	6.3 (0.7)	6.5 (0.6)	7.5 (0.6)	7.7 (0.6)	9.7 (0.5)
7	35	5.9 (1.0)	7.1 (0.8)	7.3 (0.7)	8.4 (0.6)	8.8 (0.7)	11.0 (0.5)
8	40	6.4 (1.1)	7.8 (1.0)	8.2 (0.6)	9.4 (0.8)	9.9 (0.7)	12.4 (0.5)
9	45	7.0 (1.1)	8.4 (0.9)	9.1 (0.7)	10.3 (0.7)	11.1 (0.7)	13.8 (0.6)
10	50	7.4 (1.3)	9.2 (0.9)	9.9 (0.7)	11.2 (0.7)	12.3 (0.7)	15.0 (0.5)

Table 3: Average number (and standard error) of common factors indicated by the Tracy-Widom procedure and parallel analysis (over 100 data sets), for the condition with a very small minor part ($w_{mi} = \sqrt{0.10}$) and small minor part ($w_{mi} = \sqrt{0.35}$); Q_{ma} is the number of major factors, p the number of variables.

COUNTRY	PA-random	PA-permute	TW
Australia	13	2	4
Austria	14	3	3
Bangladesh	14	2	3
Belgium	14	2	3
Brazil	14	3	3
Bulgaria	14	2	2
Canada	14	3	3
Cameroon	14	2	3
Chile	14	2	2
China	13	2	3
Colombia	9	0	3
Croatia	13	2	3
Cyprus	14	2	2
Egypt	14	2	2
Georgia	14	2	4
Germany	13	2	3
Ghana	14	2	3
Greece	14	2	2
Hong Kong	14	2	3
Hungary	14	2	2
India	13	2	3
Indonesia	12	2	3
Iran	14	2	4
Italy	14	2	3
Japan	12	2	2
Kuwait	12	2	3
Malaysia	13	0	3
Mexico	14	2	3
Nepal	10	3	3
Nigeria	14	2	2
Philippines	14	2	3
Poland	14	2	2
Portugal	14	2	2
Russia	14	2	3
Singapore	14	2	2
Slovakia	14	2	2
Slovenia	14	2	2
South Africa	14	2	2
South Korea	13	3	3
Spain	13	2	3
Switzerland	14	2	3
Thailand	14	2	3
The Netherlands	14	2	3
Turkey	14	2	3
Uganda	10	2	3
United States	14	2	2
Venezuela	12	0	2
Zimbabwe	14	2	2

Table 4: Number of estimated components according the Tracy-Widom procedure (TW), parallel analysis-random (PA-random) and parallel analysis-permute (PA-permute) per country, of the International College Survey 2001 data on universal intracultural and intercultural dimensions of the recalled frequency of emotional experience. For PA-permute, 10^4 permutations were used.

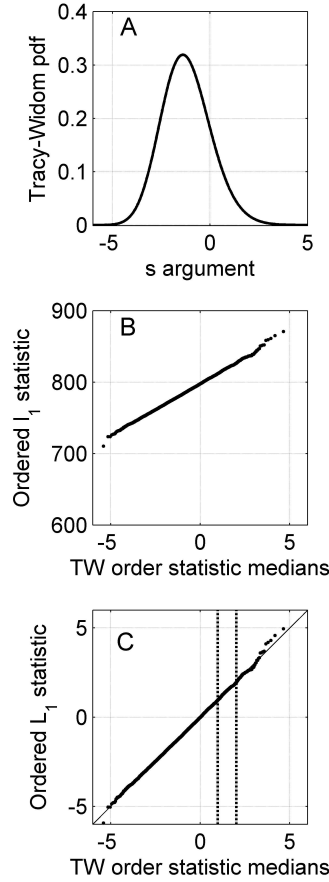


Figure 1: The Tracy-Widom probability density function (panel A). Graphical illustrations of Johnstone's theorem (panels B and C) showing the Probability-probability plots for $N = 10^4$ realizations of l_1 from sample covariance matrices of $n \times p$ ($n = p = 200$) matrices whose entries are *i.i.d.* $\mathcal{N}(0, 1)$, with the x -axis referring to the order statistic medians for the Tracy-Widom distribution; the y -axis refers to the ordered response values for the largest eigenvalue l_1 statistic (distribution of the sample eigenvalues before normalization using Equation (5)) (panel B), and the L_1 statistic (distribution of the sample eigenvalues after normalization) (panel C). Note the shift in the x - and y -axes from panel B to C. The vertical lines indicate the 95th and 99th percentiles. The order statistic medians are calculated as $F_1^{-1}((i - .5)/N)$ for $i = 1, 2, \dots, N$ where N is the number of random realizations.

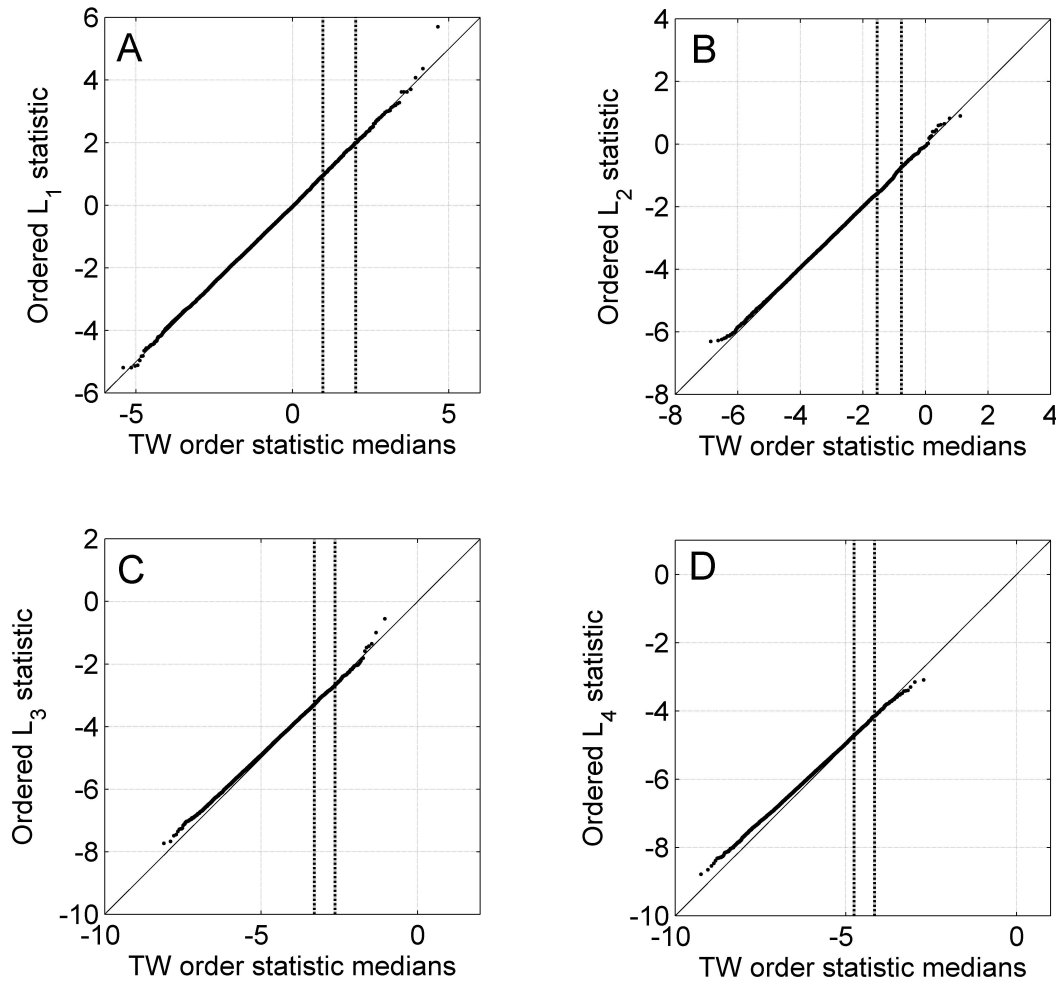


Figure 2: Probability-probability plots for $N = 10^4$ realizations of L_k (with $k = 1$ to 4 , panels A to D) from sample covariance matrices of $n \times p$ ($n = p = 300$) matrices. The vertical lines indicate the 95th and 99th percentiles. The x -axis is the order statistic medians for the Tracy-Widom distribution and the y -axis is the ordered response values for the L_k statistic.

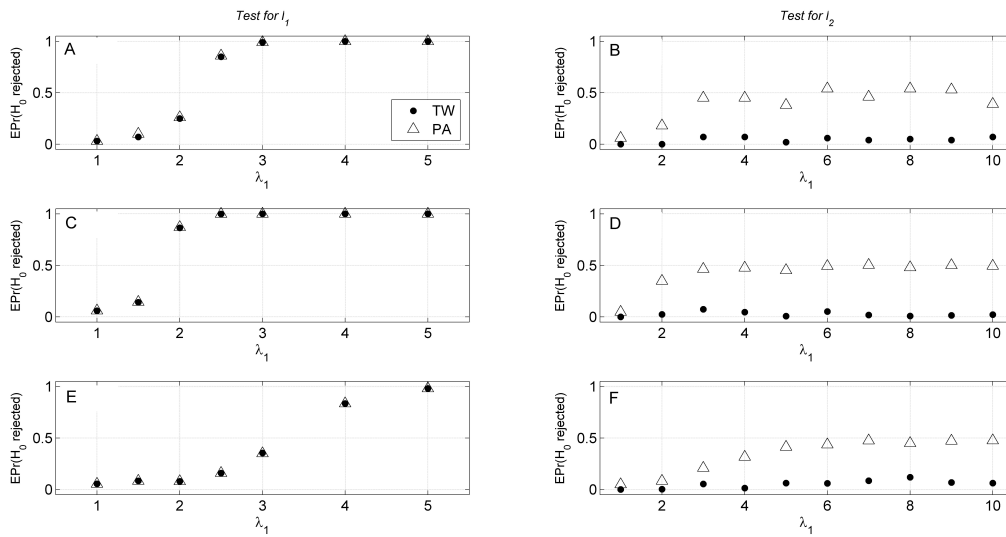


Figure 3: Null-model condition ($\Sigma = \Sigma_A$): Comparison of the performance of parallel analysis (PA) and the Tracy-Widom procedure (TW) for squared matrices 300×300 (panels A and B), rectangular matrices $n > p$ 300×100 (panels C and D) and rectangular matrices $n < p$ 100×300 (panels E and F). Successive panels give the empirical probability of rejecting H_0 for the l_1 and l_2 , respectively. Nominal $\alpha = 0.05$

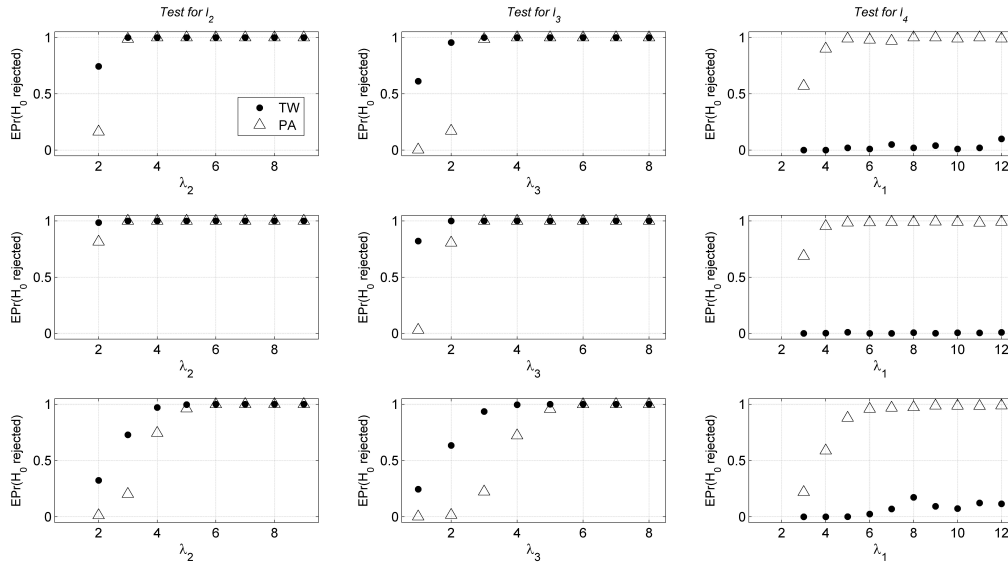


Figure 4: Spiked-model condition ($\Sigma = \Sigma_B$): Comparison of the performance of parallel analysis and the Tracy-Widom procedure for squared matrices 300×300 (panels A, B and C), rectangular matrices $n > p$ 300×100 (panels D, E and F) and rectangular matrices $n < p$ 100×300 (panels G, H and I). Successive panels (e.g., A, B, C) give the empirical probability of rejecting H_0 for the l_2 , l_3 and l_4 , respectively. Nominal $\alpha = 0.05$

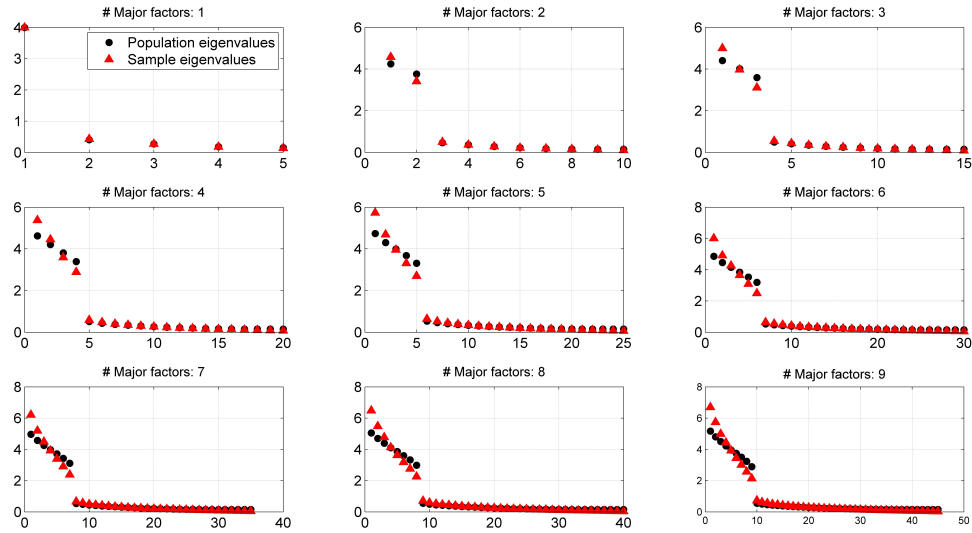


Figure 5: Population and sample eigenvalues (averaged over 100 replications) for data sets 1 to 9 in the simulation proposed by Timmerman and Lorenzo-Seva (2011).

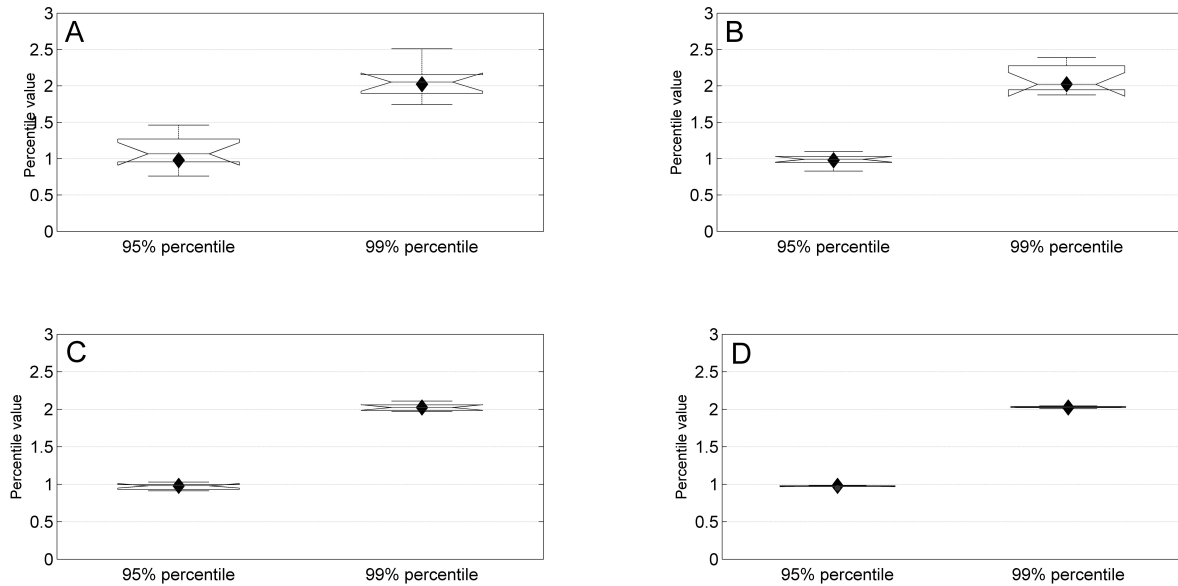


Figure 6: Variability on the estimation of the 95 and 99th percentiles of the Tracy-Widom distribution using a parallel analysis approach. The diamond symbol indicates the true percentile values. The sample size is $N = 10^q$, with $q = 2, 3, 4$ and 5 in panels A to D, respectively.